

Uncertainty Bounds for Spectral Estimation

Johan Karlsson and Tryphon T. Georgiou

Abstract—The purpose of this paper is to study metrics suitable for assessing uncertainty of power spectra when these are based on finite second-order statistics. The family of power spectra which is consistent with a given range of values for the estimated statistics represents the *uncertainty set* about the “true” power spectrum. Our aim is to quantify the size of this uncertainty set using suitable notions of distance, and in particular, to compute the diameter of the set since this represents an upper bound on the distance between any choice of a nominal element in the set and the “true” power spectrum. Since the uncertainty set may contain power spectra with lines and discontinuities, it is natural to quantify distances in the weak topology—the topology defined by continuity of moments. We provide examples of such weakly-continuous metrics and focus on particular metrics for which we can explicitly quantify spectral uncertainty. We then consider certain high resolution techniques which utilize filter-banks for pre-processing, and compute worst-case *a priori* uncertainty bounds solely on the basis of the filter dynamics. This allows the *a priori* tuning of the filter-banks for improved resolution over selected frequency bands.

Index Terms—Robust spectral estimation, uncertainty set, spectral distances, geometry of spectral measures, THREE filter design.

I. INTRODUCTION

IN practice, the estimation of power spectra in stationary time-series often relies on second-order statistics. The premise is that these are moments of an underlying power spectral distribution—the true power spectrum. Thus, the question arises as to how much is “knowable” about the distribution of power in the spectrum from such statistics.

Asymptotically, as more data accrue the convergence is guaranteed in a suitable sense, but the practical question remains on how to bound the error when only limited information is available. To this end, it is important to consider how a finite set of statistics localizes the power spectrum. Traditionally, for many applications, one relies on a particular power spectrum selected out of a variety of methods that lead to specific choices, all consistent (in different ways) with the recorded data and the estimated moments. Historically, Burg’s algorithm and the maximum entropy spectrum, and the Pisarenko harmonic decomposition are specific such choices [26], [48] and so are the correlogram and the periodogram. Thus, in general, there exists a large family of admissible power spectra which are all consistent. Bounding the values of admissible spectral density functions is an ill-posed problem (see Section IV). Instead, the natural way to quantify power

spectral uncertainty is by bounding the power on (measurable) subsets of the frequency band. Therefore, the goal of this paper is to consider the appropriate topology—the so-called weak topology, and to develop suitable metrics that can be used to quantify and measure power spectral uncertainty.

Throughout, we consider stochastic processes $\{y_t : t \in \mathbb{Z}\}$ which are discrete-time, zero-mean, and second-order stationary. A typical set of statistics for a stationary stochastic process is a finite set of covariance (or, autocorrelation) samples. The covariance samples

$$c_k := \mathcal{E}\{y_t \bar{y}_{t-k}\}, \text{ for } k = 0, \pm 1, \pm 2, \dots, \pm n,$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator, provide moment constraints for the power spectrum $d\mu$ of the process:

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\theta} d\mu(\theta) \text{ for } k = 0, \pm 1, \pm 2, \dots, \pm n. \quad (1)$$

The power spectrum is thought of as a non-negative measure on the unit circle $\mathbb{T} = \{z = e^{i\theta} : \theta \in (-\pi, \pi]\}$ (for notational simplicity also identified with the interval $(-\pi, \pi]$). We use the symbol \mathfrak{M} to denote the class of such measures and the problem of determining $d\mu \in \mathfrak{M}$ from the covariance samples (finitely or infinitely many) is known as the trigonometric moment problem. Classical theory on this problem originates in the work of Toeplitz and Carathéodory at the turn of the 20th century and has evolved into a rather deep chapter of functional analysis and of operator theory [1], [35], [22], [12], [4]. The classical monograph by Geronimus [22] contains a wide range of results on the trigonometric moment problem, the asymptotic behavior of solutions, spectral factors and optimal predictors, as well as explicit expressions for spectral envelopes [22, Theorem 5.7] (c.f. [8], [26], [14]). A more general form in which statistics may be available is when these represent the state covariance, or the output covariance, of a dynamical system driven by the stochastic process of interest. Such a dynamical system may represent a model of physical processing (bandpass filtering at sensor locations, losses, structure of sensor array, etc.) or of virtual processing (software-based) of the original time-series. Either way, covariance statistics represent (generalized) moments of the power spectrum and a theory which is completely analogous to the theory of the trigonometric moment problem is available and provides similar conclusions, see [6], [7], [14], [15], [16]. In fact, the use of generalized statistics, which relates to beamspace processing, was explored in [7], [15] as a way to improve resolution in power spectral estimation over selected frequency bands. More recent work addresses spectral estimation with priors, computational issues, as well as important multivariate generalizations [3], [5], [9], [10], [11], [17], [18], [20], [21], [38], [40], [43], [44].

The framework of the present work involves such moment problems specified by covariance statistics. Invariably, moment statistics are estimated from a finite observation record and are

Supported by Swedish Research Council, Göran Gustafsson Foundation, National Science Foundation Grant No. CCF-1218388 and Grant No. 1027696, Air Force Office of Scientific Research under Grant FA9550-12-1-0319, and the Vincentine Hermes-Luh Endowment.

Johan Karlsson is with the department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611, jkarlsson@ufl.edu. Tryphon T. Georgiou is with the Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA, tryphon@ece.umn.edu

known with limited accuracy. Thus, in a typical experiment, as the observation record of a time-series increases so does the accuracy and the length of the estimated partial covariance sequence. Our goal is to develop metrics that can be used to quantify spectral uncertainty. More specifically, phrased in the context of the trigonometric moment problem, we seek metrics between power spectra that have the following properties:

- (i) given a finite set of covariance samples, the family of consistent power spectra has a finite diameter, and
- (ii) the diameter of the uncertain set of power spectra shrinks to zero as both, the accuracy of the covariance samples increases and their number tends to infinity.

The latter condition is dictated by the fact that the trigonometric moment problem is known to be determined, i.e., there is a unique power spectrum which is consistent with an infinite sequence of covariances. As we will explain below (in Section III), the proper topology which allows for these properties to hold is the weak topology on measures (cf., [27, page 8]). There is a variety of metrics that can be used to metrize this topology, and thus, in principle, to quantify spectral uncertainty. A contribution of this work is to suggest a class of metrics for which the radius of spectral uncertainty and *a priori bounds* are computable given a finite set of (error-free) statistics.

In Section II we review the trigonometric moment problem and relevant concepts in functional analysis. In Section III we define power-spectral uncertainty sets and discuss the relevance of weakly continuous metrics. In Section IV we present a collection of weakly continuous metrics that, in different ways, are suitable for metrizing the space of power spectra. In Section V we compute the diameter of uncertainty sets, for a particular choice of a metric, and elaborate on the limit properties of this uncertainty quantification. In Section VI we present an example that elucidates the relevance and applicability of the results in practice. In Section VII we explain how the framework applies in the context of generalized statistics. In Section VIII we highlight the use of this quantification of uncertainty in filter design—we show how to tune a filter-bank so as limit spectral uncertainty over some frequency range of interest. In the concluding section (Section IX) we summarize the results and outline possible future directions.

II. THE TRIGONOMETRIC MOMENT PROBLEM, SPECTRAL REPRESENTATIONS, AND WEAK CONVERGENCE

The covariances c_k , $k = 0, \pm 1, \pm 2, \dots$, of a stationary random process $\{y_t : t \in \mathbb{Z}\}$ are the Fourier coefficients of the spectral measure $d\mu$ as in (1). These are characterized by the non-negativity of the Toeplitz matrices [25], [26]

$$T_n = \begin{bmatrix} c_0 & c_{-1} & \cdots & c_{-n} \\ c_1 & c_0 & \cdots & c_{-n+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_n & c_{n-1} & \cdots & c_0 \end{bmatrix},$$

for $n = 0, 1, \dots$. When $T_n > 0$ for $n \leq k$ and singular for $n = k + 1$, then it is also singular for all $n > k$ and $\text{rank}(T_{k+\ell}) = \text{rank}(T_k) = k + 1$ for all $\ell \geq 1$. In this case, $d\mu$ is singular with respect to the Lebesgue measure and consists of finitely many “spectral lines,” equal in number to $\text{rank}(T_n)$

[25, page 148]. Because $d\mu$ is a real measure, $c_k = \bar{c}_{-k}$ for $k = 0, 1, \dots$, hence we use only positive indices and refer by

$$\mathbf{c}_{0:n} := (c_0, c_1, \dots, c_n)$$

to the vector of the first $(n + 1)$ moments, and by

$$\mathbf{c} := (c_0, c_1, \dots)$$

to the infinite sequence. The sequence \mathbf{c} is said to be *positive* if $T_n > 0$ for all n . Similarly $\mathbf{c}_{0:n}$ is said to be *positive* if $T_n > 0$. Accordingly, the term *non-negative* is used when the relevant Toeplitz matrices are non-negative definite.

As noted in the introduction, the power spectrum of a discrete-time stationary process is a bounded non-negative measure on the unit circle. The derivative (of its absolutely continuous part) is referred to as the spectral density function, while the singular part typically contains jumps (spectral lines) associated with the presence of sinusoidal components. In general, the singular part may have a more complicated mathematical structure that allocates “energy” on a set of measure zero without the need for distinct spectral lines [25, page 5]. From a mathematical viewpoint such spectra are important as they represent limits of more palatable spectra, and hence, represent a form of completion.

The natural topology where such limits ought to be considered is the so-called *weak topology*. This topology is also known as the weak* topology in functional analysis—a term which is less frequently used in the context of measures. The weak topology is defined in terms of convergence of linear functionals and is explained next. We denote by $C(\mathbb{T})$ the class of real-valued continuous functions on \mathbb{T} . It is quite standard that the space of bounded linear functionals $\Lambda : C(\mathbb{T}) \rightarrow \mathbb{R}$, can be identified with the space of bounded measures on \mathbb{T} [27, page 7]. More specifically, any bounded functional Λ can be represented in the form

$$\Lambda(f) = \int_{\mathbb{T}} f(t) d\mu(t) \text{ for all } f \in C(\mathbb{T}),$$

with $d\mu$ being the corresponding measure—this is the Riesz representation theorem. Continuous functions now serve as “test functions” to differentiate between measures. Bounds on the corresponding integrals define the weak topology: a sequence of measures $d\mu_n$, $n = 1, 2, \dots$, converges to $d\mu$ in the weak topology if $\int f d\mu_n \rightarrow \int f d\mu$ for every $f \in C(\mathbb{T})$. Thus, for any two measures that are different, there exists a continuous function that the two measures integrate to different values. In this setting, a measure can be specified uniquely by its Fourier coefficients. In fact, given a positive sequence \mathbf{c} , the unique corresponding measure $d\mu$ can be determined as the limit in the weak topology of finite Fourier sums or Cesaro means [27, page 24].

Non-negative measures are naturally associated with analytic and harmonic functions—a connection which has been exploited in classical circuit theory in the context of passivity. Herglotz’ theorem [1] states that if $d\mu$ is a bounded non-negative measure on \mathbb{T} , then

$$H[d\mu](z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) \quad (2)$$

is analytic in $\mathbb{D} := \{z : |z| < 1\}$ and the real part is non-negative. Such functions are referred to as either “positive-real” or, as Carathéodory functions. Conversely, any positive-real function can be represented (modulo an imaginary constant) by the above formula for a suitable non-negative measure. The Poisson integral of a non-negative measure $d\mu$

$$P[d\mu](z) := \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(t - \theta) d\mu(\theta), \quad z = re^{it}, \quad (3)$$

where $P_r(\theta) = \frac{1-r^2}{|1-re^{i\theta}|^2}$ is the Poisson kernel, is a harmonic function which is non-negative in \mathbb{D} and is equal to the real part of $H[d\mu](z)$. Given either a positive-real function $H(z)$, or its real part $P(z)$, the measure $d\mu$ such that $H(z) = H[d\mu](z)$ and $P(z) = P[d\mu](z)$ is uniquely determined by the limit of $P(re^{i\theta})d\theta \rightarrow d\mu$ as $r \rightarrow 1$ in the weak topology [27, page 33]. Thus, power spectra are, in a very precise sense, boundary limits of the (harmonic) real parts of positive-real functions.

III. UNCERTAINTY OF SPECTRAL ESTIMATES

We postulate a situation where covariances $\mathbf{c}_{0:n}$ are estimated from sample of a stochastic process $\{y_t\}_{t \in \mathbb{Z}}$ with power spectrum $d\nu$, and where the estimation error in the entries of $\mathbf{c}_{0:n}$ are bounded by ϵ^1 . Thus, the “true” spectrum $d\nu$ belongs to the uncertainty set

$$\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon} := \left\{ d\mu \geq 0 : \left| c_k - \int_{-\pi}^{\pi} e^{-ik\theta} d\mu \right| < \epsilon, k = 0, 1, \dots, n \right\}.$$

Likewise, any choice for a “nominal” spectrum $d\hat{\nu}$ consistent with our assumptions will also belong to $\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon}$. Therefore, the distance between the two will be bounded by the diameter of the uncertainty set,

$$\rho_{\delta}(\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon}) := \sup\{\delta(d\mu_0, d\mu_1) : d\mu_0, d\mu_1 \in \mathcal{F}_{\mathbf{c}_{0:n}, \epsilon}\},$$

where δ is a suitable metric at hand. Thus, our goal in this paper is to seek metrics δ on the space of positive measures \mathfrak{M} that provide a meaningful and computationally tractable notion of a diameter for $\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon}$ thereby quantifying modeling uncertainty in the spectral domain. To narrow down the search for suitable metrics, consider the scenario when the length of the data increases, and hence the accuracy as well as the number of covariance lags increases. In the limit, as the estimation error goes to zero and the number n of covariance lags goes to infinity, the uncertainty set shrinks to the singleton

$$\{d\nu\} = \bigcap_{n \in \mathbb{N}} \mathcal{F}_{\mathbf{c}_{0:n}, \epsilon_n}, \quad (\epsilon_n \rightarrow 0 \text{ as } n \rightarrow \infty).$$

This is due to the fact that an infinite limit sequence \mathbf{c} defines a unique power spectrum—the trigonometric problem is determinate. The diameter should reflect this shrinkage to a singleton and tend to zero. For this to happen, the underlying metric needs to be weakly continuous as stated next.

Theorem 1: Let δ be a metric on \mathfrak{M} . Then

$$\rho_{\delta}(\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon_n}) \rightarrow 0 \text{ as } \epsilon_n \rightarrow 0 \text{ and } n \rightarrow \infty, \quad (4)$$

for every covariance sequence \mathbf{c} if and only if δ is weakly continuous.

¹The more realistic situation, where the confidence intervals degrade with the order of covariance lags, can be dealt with in a similar manner, albeit with a bit more cumbersome notation.

Proof: This can be seen by comparing the definition of $\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon_n}$ with the definition of open sets in the weak topology. See the appendix for a detailed proof. ■

Remark 2: Occasionally one may have additional a priori knowledge on the structure and smoothness of the power spectrum which would further limit the uncertainty set. Quantifying such “structured” uncertainty would necessarily be problem-specific and is not considered in the present work. Instead, we take a viewpoint that allows comparing power spectra in a unified way, regardless smoothness, presence of spectral lines, or membership in a specific class of models. □

We now consider the case where the finite covariance sample $\mathbf{c}_{0:n}$ is known exactly. If $\mathbf{c}_{0:n}$ is positive, then the uncertainty set

$$\mathcal{F}_{\mathbf{c}_{0:n}} := \left\{ d\mu \geq 0 : c_k = \int_{-\pi}^{\pi} e^{-ik\theta} d\mu, k = 0, 1, \dots, n \right\}$$

contains infinitely many power spectra. If $\mathbf{c}_{0:n}$ is only non-negative, and hence T_n is singular, then the family $\mathcal{F}_{\mathbf{c}_{0:n}}$ consists of the single power spectrum $d\nu$ [25, page 148]. The following two results are immediate corollaries of Theorem 1. The first one treats the case where the number of covariance lags goes to infinity, while the second, treats the case where the values of the covariance lags tend to those of a singular sequence. In both cases the diameter of the uncertainty set necessarily goes to zero for a weakly continuous metric.

Corollary 3: Let \mathbf{c} be a non-negative sequence and let δ be a weakly continuous metric. Then

$$\rho_{\delta}(\mathcal{F}_{\mathbf{c}_{0:n}}) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Proof: This follows directly from Theorem 1 and by noting that

$$\mathcal{F}_{\mathbf{c}_{0:n}} \subset \mathcal{F}_{\mathbf{c}_{0:n}, \epsilon}$$

for any $\epsilon > 0$. It also follows from [22, §1.16] in view of Proposition 10 in Section IV-C below. ■

Corollary 4: Let $\mathbf{c}_{0:n}$ be a vector of covariance lags such that the corresponding T_n is a singular Toeplitz matrix, and let $\hat{\mathbf{c}}_{0:n}(k)$ ($k = 1, 2, \dots$) be a sequence of vectors of covariance lags tending to $\mathbf{c}_{0:n}$. If δ is a weakly continuous metric then

$$\rho_{\delta}(\mathcal{F}_{\hat{\mathbf{c}}_{0:n}(k)}) \rightarrow 0, \text{ as } k \rightarrow \infty.$$

Proof: Follows directly from Theorem 1. See also [31] for an independent detailed argument. ■

Remark 5: It should be noted that the total variation ($\int |d\mu_0 - d\mu_1|$) is not weakly continuous and therefore the conclusions of the two corollaries would fail if this was used as the metric. To see this, note that if $\mathbf{c}_{0:n}$ is positive, then $\mathcal{F}_{\mathbf{c}_{0:n}}$ contains infinitely many measures and among them at least two singular measures with non-overlapping support, i.e., $\text{supp}(d\mu_0) \cap \text{supp}(d\mu_1) = \emptyset$ (e.g., see [35]). Then the total variation of their difference is always $2c_0$. □

IV. WEAKLY CONTINUOUS METRICS

In general, a finite set of second-order statistics cannot dictate the precise value of the power spectrum locally. Indeed, given any finite positive sequence $\mathbf{c}_{0:n}$ and any $\theta_0 \in (-\pi, \pi]$, then for any value $\alpha \geq 0$ there exists an $\epsilon > 0$ and an absolutely continuous measure $d\mu = f d\theta \in \mathcal{F}_{\mathbf{c}_{0:n}}$ such that

$$f(\theta) = \alpha \text{ for } \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon).$$

What can be said instead, is that the range of values

$$\left\{ \int_{\mathbb{T}} g d\mu : d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}} \right\}, \quad (5)$$

for any particular test function $g \in C(\mathbb{T})$, is bounded. Furthermore, as $n \rightarrow \infty$, this range tends to zero. In fact, due to weak continuity, the range of values tend to zero for any of the scenarios in Theorem 1 and its two corollaries. Finding the maximum and the minimum of (5) is a linear programming problem on an infinite dimensional domain. Provided g is symmetric real and the covariance sequence $\mathbf{c}_{0:n}$ is real, the dual problems, which give the lower and upper bounds of (5), are

$$\max \left\{ \lambda \mathbf{c}_{0:n}^T : \sum_{k=0}^n \lambda_k \cos(k\theta) \leq g(\theta), \theta \in (-\pi, \pi] \right\}, \quad (6)$$

$$\min \left\{ \lambda \mathbf{c}_{0:n}^T : g(\theta) \leq \sum_{k=0}^n \lambda_k \cos(k\theta), \theta \in (-\pi, \pi] \right\}, \quad (7)$$

where $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_n)$ are Lagrange multipliers.

Remark 6: Along these lines Lang and Marzetta in [36], [37] sought to quantify the maximal and minimal spectral mass in a specified interval given the covariances $\mathbf{c}_{0:n}$. To this end we may take $g = \chi_I$ the characteristic function of an interval I , that is, $\chi_I(\theta) = 1$ if $\theta \in I$ and 0 otherwise. Lower and upper bounds on $\int_I d\mu$ are finite and are then given by (6) and (7), respectively. However, since $g = \chi_I$ is not continuous, the mass in an interval is not a weakly continuous quantity, and the requirements in Corollary 3 does not hold. In fact, for this case the gap between the upper and lower bound does not necessarily converge to zero as n goes to infinity. This occurs, e.g., in the case when the true spectrum has a spectral line at an end point of the interval. \square

A class of weakly continuous metrics can be sought in the form

$$\delta(d\mu_0, d\mu_1) = \sup_{\xi \in K} \left| \int_{\mathbb{T}} g_{\xi}(d\mu_0 - d\mu_1) \right|, \quad (8)$$

for $\{g_{\xi}\}_{\xi \in K} \subset C(\mathbb{T})$, provided the family $\{g_{\xi}\}_{\xi \in K}$ of test functions is sufficiently rich to distinguish between measures and yet, small enough so that continuity is ensured. The precise conditions are given next.

Proposition 7: The functional $\delta(d\mu_0, d\mu_1)$ defined in (8) is a weakly continuous metric if and only if the following two conditions hold:

- (a) for any two measures $d\mu_0, d\mu_1 \in \mathfrak{M}$, there is a $\xi \in K$ such that $\int_{\mathbb{T}} g_{\xi} d\mu_0 \neq \int_{\mathbb{T}} g_{\xi} d\mu_1$, and
- (b) the set $\{g_{\xi}\}_{\xi \in K}$ in $C(\mathbb{T})$ is equicontinuous² and uniformly bounded.

Proof: See the appendix. \blacksquare

In essence, condition (a) ensures positivity while condition (b) ensures weak continuity. The triangle inequality and symmetry always hold for such δ . The total variation norm is an example of why (b) is needed—it is a norm of the form (8) where the set of test function are the $C(\mathbb{T})$ unit ball, $\{g : \|g\|_{\infty} \leq 1\}$, but it is not weakly continuous. This is due to the fact that the unit ball in $C(\mathbb{T})$ is not equicontinuous.

²A family of functions $\{g_{\xi}\}_{\xi \in K} \subset C(\mathbb{T})$ is said to be equicontinuous if for any ϵ there exists a γ such that $|g_{\xi}(\theta_1) - g_{\xi}(\theta_2)| < \epsilon$ if $|\theta_1 - \theta_2| < \gamma$ for all $\theta_1, \theta_2 \in \mathbb{T}$, and $\xi \in K$.

Remark 8: A more general family of distances are of the form

$$\delta(d\mu_0, d\mu_1) = \sup_{\substack{g_0(\theta) \in K_0, g_1(\phi) \in K_1, \\ g_0(\theta) + g_1(\phi) \in K}} \int_{\mathbb{T}} g_0 d\mu_0 + \int_{\mathbb{T}} g_1 d\mu_1$$

where $K_0, K_1 \subset C(\mathbb{T})$ and $K \subset C(\mathbb{T} \times \mathbb{T})$. By selecting the sets K_0, K_1 , and K properly, δ (or a monotone function of δ) will be a weakly continuous metric. One such example is the metrics based on optimal transportation treated in [19], where the metrics have non-local properties such as geodesics which preserve lumpedness. \square

Next we consider three ways for devising weakly continuous metrics. The first uses smoothing of power spectra to be compared by suitable test functions in a way that is analogous to the use of classical window kernels in periodogram estimation [48]. The second is based on Monge-Kantorovich optimal mass transportation where a cost is associated with mismatch in the frequency range where power resides. In this geometry, optimal-transport geodesics may be used to model slow time-varying drift in the spectral power of non-stationary time-series [19] — such models for non-stationarity lessen artifacts present when using ordinary interpolation (e.g., fade-in fade-out [30]). The third is based on Poisson kernels and is more suitable for differentiating spectra based on their content on specified frequency bands. The connection between Poisson kernels and the analytic and harmonic functions in (2) and (3) allows for evaluating bounds and the diameter of the uncertainty set with respect to the corresponding distances. This will be explored in the case where finitely many error-free covariances are known in Sections V to VIII.

A. Metrics based on smoothing

A simple way to devise weakly continuous metrics which has a classical flavor is to first smoothing the measures via convolution with a fixed suitable continuous function, and then to compare the smoothed spectral densities. This echoes the use of windowing Fourier techniques in the time domain [48] where a suitable choice of a window is used to trade-off resolution and variance of the estimator. Likewise here, the choice of a windowing function determines the resolution of the metric.

Thus, let $g \in C(\mathbb{T})$ be such a windowing function, and define

$$\delta_{\text{smooth},g}(d\mu_0, d\mu_1) := \|g * (d\mu_0 - d\mu_1)\|_{\infty}.$$

Here,

$$(g * d\mu)(\xi) = \int_{-\pi}^{\pi} g(\xi - \theta) d\mu(\theta)$$

denotes the circular convolution and $\|\cdot\|_{\infty}$ the L_{∞} norm. In the view of Proposition 7, $\delta_{\text{smooth},g}$ is of the form

$$\|g * (d\mu_0 - d\mu_1)\|_{\infty} = \sup_{\xi \in (-\pi, \pi]} \left| \int_{-\pi}^{\pi} g(\xi - \theta) (d\mu_0(\theta) - d\mu_1(\theta)) \right|,$$

and hence, condition (b) of the proposition holds. In addition, the chosen convolution-kernel functions must not have any zero Fourier coefficient, otherwise the approach will fail to differentiate between certain measures. To see this, let $g(\theta) =$

$\sum_{k=-\infty}^{\infty} g_k e^{ik\theta}$ and let $(\dots, a_{-1}, a_0, a_1, \dots)$ be the Fourier coefficients of $d\mu_0(\theta) - d\mu_1(\theta)$, then

$$g * (d\mu_0 - d\mu_1)(\xi) = \sum_{k=-\infty}^{\infty} g_{-k} a_k e^{ik\xi}.$$

If $g_k \neq 0$ for all $k \in \mathbb{Z}$, the above expression cannot vanish identically unless all the a_k 's are zero, in which case $d\mu_0 = d\mu_1$. In this case (a) holds and it follows from Proposition 7 that $\delta_{\text{smooth},g}(d\mu_0, d\mu_1)$ is a weakly continuous metric. This leads to the next proposition.

Proposition 9: Let $g \in C(\mathbb{T})$ be a windowing function with non-vanishing Fourier coefficients. Then $\delta_{\text{smooth},g}(d\mu_0, d\mu_1)$ is a weakly continuous metric.

B. Metrics based on optimal transportation

A rapidly growing literature [50] on a classical problem, known as the Monge-Kantorovich transportation problem, has impacted a wide range of disciplines, from probability theory to fluid dynamics and economy [42]. Optimal transportation refers to the correspondence between distributions of masses that induce the least amount of transportation cost³. The optimal transportation cost between two probability distributions induces weakly continuous metrics, known as Wasserstein metrics, which are extensively used in probability theory. In order to handle more general distributions we need a suitable modification to compare unequal masses. This we do next and connect with the formalism in (8).

The Monge-Kantorovich transportation problem amounts to minimizing the cost of transportation between two distributions of equal mass, e.g., $d\mu_0$ and $d\mu_1$ where $\int_{\mathbb{T}} d\mu_0 = \int_{\mathbb{T}} d\mu_1$. In this, a transportation plan $d\pi(\theta, \phi)$ is sought which corresponds to a non-negative distribution on $\mathbb{T} \times \mathbb{T}$ and is such that

$$\int_{\theta \in \mathbb{T}} d\pi(\theta, \phi) = d\mu_0(\phi) \text{ and } \int_{\phi \in \mathbb{T}} d\pi(\theta, \phi) = d\mu_1(\theta). \quad (9)$$

Then, the minimal cost

$$\min \left\{ \int_{\mathbb{T} \times \mathbb{T}} |\theta - \phi| d\pi(\theta, \phi) : d\pi \text{ satisfies (9)} \right\}$$

is the Wasserstein-1 distance between $d\mu_0$ and $d\mu_1$, and is a weakly continuous metric (see, e.g., [50, chapter 7]). This problem admits a dual formulation, known as the Kantorovich duality:

$$W_1(d\mu_0, d\mu_1) = \max_{\|g\|_L \leq 1} \int g(d\mu_0 - d\mu_1),$$

where $\|f\|_L = \sup_{\theta, \phi} \frac{|f(\theta) - f(\phi)|}{|\theta - \phi|}$ denotes the Lipschitz norm.

Power spectra, in general, cannot be expected to have the same total mass. In this case, $\delta_{1,\kappa}(d\mu_0, d\mu_1)$ defined by

$$\inf_{\int d\nu_0 = \int d\nu_1} W_1(d\nu_0, d\nu_1) + \kappa \sum_{i=0}^1 \int_{\mathbb{T}} |d\mu_i - d\nu_i|, \quad (10)$$

is a weakly continuous metric for an arbitrary but fixed $\kappa > 0$. The interpretation is that $d\mu_0$ and $d\mu_1$ are perturbations of the two underlying measures $d\nu_0$ and $d\nu_1$, respectively, which

have equal mass. Then, the cost of transporting $d\mu_0$ and $d\mu_1$ to one another can be thought of as the cost of transporting $d\nu_0$ and $d\nu_1$, to one another, plus the size of their respective perturbations from $d\mu_0$ and $d\mu_1$. This is introduced in [19] and this metric admits a dual formulation

$$\delta_{1,\kappa}(d\mu_0, d\mu_1) = \max_{\substack{\|g\|_{\infty} \leq \kappa \\ \|g\|_L \leq 1}} \int g(d\mu_0 - d\mu_1),$$

which is in the form of the Proposition 7. Various other generalizations of the transportation distance that apply to power spectra are also being proposed and studied in [19].

C. Metrics based on the Poisson kernel

Power spectra are weak limits of the real part of analytic functions on the unit disc, as indicated earlier. Comparison of these functions induces weakly continuous metrics which readily fall under the framework of (8). Interestingly, this approach allows for both the computation of explicit/analytic bounds on uncertainty sets (see Section V) and for specifying a frequency dependent resolution of a metric (see Remark 11 and the example in Section VII).

Recall from Section II that the harmonic function associated with a measure is the Poisson integral, defined as

$$P[d\mu](z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(t - \theta) d\mu(\theta), \quad z = re^{it}.$$

Weak convergence of measures is equivalent to certain types of convergence of their harmonic counterpart.

Proposition 10: Let $\{d\mu_k\}_{k=1}^{\infty}$ be a sequence of uniformly bounded signed measures on \mathbb{T} , let $d\mu$ be a bounded measure on \mathbb{T} , and let $u(z) = P[d\mu](z)$, $u_k(z) = P[d\mu_k](z)$ be their corresponding Poisson integrals. The following statements are equivalent:

- (a) $d\mu_k \rightarrow d\mu$ weakly,
- (b) $u_k(z) \rightarrow u(z)$ pointwise $\forall z \in \mathbb{D}$,
- (c) $u_k(z) \rightarrow u(z)$ in $L_1(\mathbb{D})$,
- (d) $u_k(z) \rightarrow u(z)$ uniformly on every compact subset of \mathbb{D} .

Proof: The proof is given in the appendix. ■

Each of the statements (b), (c), and (d) may be used for devising weakly continuous metrics. We shall focus on the statement (d), indicating that weakly continuous metrics can be constructed by comparing the harmonic functions on a subset of \mathbb{D} . In fact, the maximal distance between the harmonic functions on a closed non-finite set $K \subset \mathbb{D}$ gives rise to a weakly continuous metric

$$\delta_K(d\mu_0, d\mu_1) = \max_{z \in K} |P(d\mu_0 - d\mu_1)(z)|. \quad (11)$$

This is true, since the resulting family of the Poisson kernels satisfies the properties in Proposition 7. To see this, first note that any two harmonic functions which coincides on K , a closed non-finite set inside \mathbb{D} , must be identical, hence (a) is satisfied. Further more, since $K \subset \gamma\mathbb{D}$ for some $\gamma < 1$, the magnitude and derivative of $P_r(t - \theta)$ is uniformly bounded when $re^{it} \in K$, hence (b) holds.

Remark 11: In practice, it is often the case that one is interested in comparing spectra over selected frequency bands. To this end, various schemes have been considered which rely on pre-processing with a choice of “weighting” filters

³L. Kantorovich received the 1975 Nobel Prize for the impact of this theory on allocation of economic resources.

and filter banks (see e.g., [8], [49], and [6], [16]). The choice of the point-set K in (11) can be used to dictate the resolution of the metric over such frequency bands. To see how this can be done, consider K to designate an arc $\{\xi = re^{i\theta} : \theta \in [\theta_0 - \epsilon, \theta_0 + \epsilon]\}$. This satisfies the conditions of Proposition 7 and thus, δ_K is a weakly continuous metric. At the same time, the values $P[d\mu](\xi)$, with $\xi \in K$, represent the variance at the output of a filter with transfer function $z/(z - \xi)$. These are bandpass filters with a center frequency $\arg(\xi)$ and bandwidth which depends on the choice of r . Thus, in essence, the metric compares the respective variance after the spectra have been weighted by a continuum (for $\xi \in K$) of such frequency-selective bank of filters. \square

V. THE SIZE OF THE UNCERTAINTY SET

The diameter of the uncertainty set with respect to the distance δ_K turns out to be especially easy to compute – it is realized as the distance between two “diametrically opposite” measures with only $n + 1$ spectral lines each (i.e., measures having compact support). This is the content of the following proposition.

Proposition 12: Let $\mathbf{c}_{0:n}$ be a positive covariance sequence and let $K \subset \mathbb{D}$ be closed. Then

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}}) = \max_{z \in K} \left\{ 2 \left(\left| \frac{\frac{2}{1-z\bar{z}} + \langle b_z, d_z \rangle_{T^{-1}}}{\langle b_z, b_z \rangle_{T^{-1}}} \right|^2 - \frac{\langle d_z, d_z \rangle_{T^{-1}}}{\langle b_z, b_z \rangle_{T^{-1}}} \right)^{\frac{1}{2}} \right\},$$

where

$$b_z = \begin{pmatrix} z^{-1} \\ z^{-2} \\ \vdots \\ z^{-n-1} \end{pmatrix}, \quad d_z = \begin{pmatrix} z^{-1}(c_0) \\ z^{-2}(c_0 + 2c_1z) \\ \vdots \\ z^{-n-1}(c_0 + 2c_1z + \dots + 2c_nz^n) \end{pmatrix},$$

and $\langle x, y \rangle_{T^{-1}}$ denotes the inner product

$$\langle x, y \rangle_{T^{-1}} := y^* T_n^{-1} x.$$

Furthermore, $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ is attained as the distance between two elements of $\mathcal{F}_{\mathbf{c}_{0:n}}$ which are both singular with support containing at most $n + 1$ points.

Proof: The proof is given in the appendix. \blacksquare

Both claims in Proposition 12 can be used separately for computing $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$. The first one suggests finding a maximum of a real-valued function over K . The second claim suggests a search for a maximum of $\delta_K(d\mu_1, d\mu_2)$ over a rather small subset of $\text{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$, namely nonnegative sequences $\mathbf{c}_{0:(n+1)}$ parametrized by c_{n+1} ; i.e., solutions of the quadratic equation

$$\det(T_{n+1}) = 0. \quad (12)$$

The (complex) values for c_{n+1} satisfying (12), lie on a circle in the complex plane, and hence, computation of $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ requires search on a torus (each of the two extremal $d\mu_1, d\mu_2$ where the diameter is attained can be thought of as points on the circle).

We elucidate this with an example. Figure 1 shows $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ for

$$\mathbf{c}_{0:2} = (1, c_1, c_2)$$

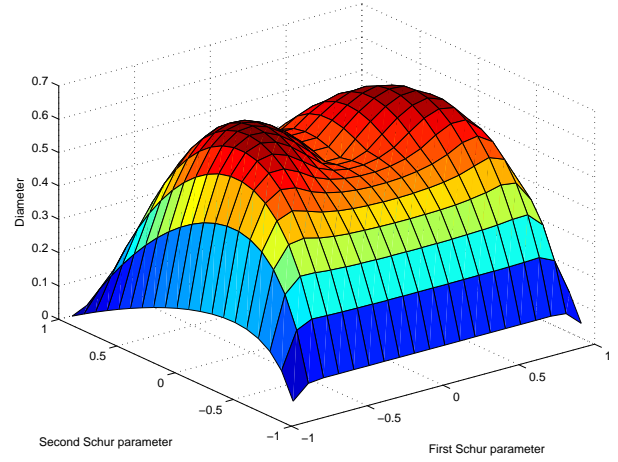


Fig. 1. The uncertainty diameter ρ_{δ_K} as a function of γ_1, γ_2 when $c_0 = 1$ and $K = \{z : |z| \leq 0.5\}$.

as a function of the corresponding *partial autocorrelation coefficients*, also known as *Schur parameters* (see the appendix),

$$\begin{aligned} -1 < \gamma_1 &:= c_1 < 1, \\ -1 \leq \gamma_2 &:= \frac{\det \begin{pmatrix} c_1 & c_2 \\ 1 & c_1 \end{pmatrix}}{\det \begin{pmatrix} 1 & c_1 \\ \bar{c}_1 & 1 \end{pmatrix}} \leq 1, \end{aligned}$$

and K is taken as $\{z : |z| \leq 0.5\} \subset \mathbb{D}$.

The plot confirms that the diameter decreases to zero as the parameters or, alternatively, the covariances c_1 and c_2 , tend to the boundary of the “positive” region (which in the Schur coordinates corresponds to the unit square). However, it is interesting to note that the diameter of $\mathcal{F}_{\mathbf{c}_{0:n}}$ as a function of $\mathbf{c}_{0:n}$ has several local maxima. This maximal diameter may be explicitly calculated, hence provides an *a priori* bound on the uncertainty.

Theorem 13: Let $r = \max(|z| : z \in K)$. Then

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}}) \leq \frac{4c_0|r|^{n+1}}{1 - |r|^2}. \quad (13)$$

Further, (13) holds with equality if and only if $\mathbf{c}_{0:n} = (c_0, c_0\bar{\alpha}, c_0\bar{\alpha}^2, \dots, c_0\bar{\alpha}^n)$ for some $\alpha \in K$ with $|\alpha| = r$.

Proof: The proof is given in the appendix. \blacksquare

Remark 14: Computation of the diameter $\rho_{\delta}(\mathcal{F}_{\mathbf{c}_{0:n}})$ of the uncertainty set amounts to solving the infinite-dimensional optimization problem

$$\sup\{\delta(d\mu_1, d\mu_2) : d\mu_1, d\mu_2 \in \mathcal{F}_{\mathbf{c}_{0:n}}\}. \quad (14)$$

If δ is a weakly continuous and jointly convex function, then the diameter is attained as the precise distance between two elements which are extreme points $\mathcal{F}_{\mathbf{c}_{0:n}}$. Extreme points are the points with the property that they themselves are not a convex combination of other elements in the set; the set of extreme points is denoted by $\text{ext}(\cdot)$. Then, $d\mu \in \text{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$ if and only if $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$ and the support of $d\mu$ consists of at most $2n + 1$ points (see [31]). Thus, $\text{ext}(\mathcal{F}_{\mathbf{c}_{0:n}})$ admits a finite dimensional characterization and (14) reduces to a finite dimensional problem. \square

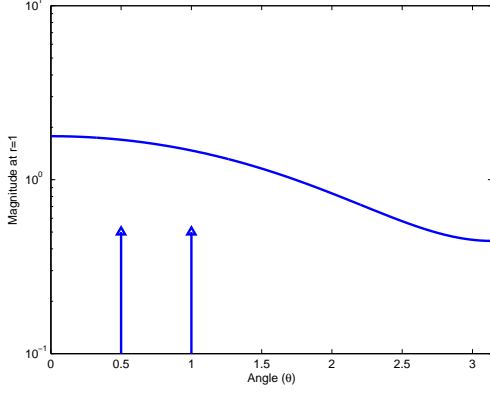


Fig. 2. The “true” power spectrum $d\nu$.

VI. IDENTIFICATION IN A WEAK SENSE

In this section we elucidate how the uncertainty set is affected by the number of moments and show that spectra may be close in the weak sense even though they are qualitatively very different.

Consider the stochastic process

$$y_t = \cos(0.5t + \varphi_1) + \cos(t + \varphi_2) + w_t + \frac{1}{3}w_{t-1}$$

where w_t is a white noise process and φ_1, φ_2 are random variables with uniform distribution on $(-\pi, \pi]$. The power spectrum $d\nu$ is depicted in Figure 2 and the spectrum has both an absolutely continuous part as well as a singular part. We would like to identify this spectrum relying on covariance data and derive bounds on the estimation error. We will use the metric δ_K where $K = \{z : |z| = 0.9\}$, i.e.,

$$\delta_K(d\mu_0, d\mu_1) = \sup_{|z|=0.9} |P(d\mu_0 - d\mu_1)(z)|.$$

Let \mathbf{c} be the covariance sequence of $d\nu$ and let $d\mu_5$ and $d\mu_{20}$ be the power spectra with highest entropy in the sets $\mathcal{F}_{\mathbf{c}_{0:5}}$ and $\mathcal{F}_{\mathbf{c}_{0:20}}$, respectively. Figure 3 compares $d\mu_5$ and $d\nu$ where the estimation error and the uncertainty diameter are

$$\delta_K(d\nu, d\mu_5) = 5.66, \quad \rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:5}}) = 20.79.$$

The first subplot shows and compares these two power spectra. The second subplot displays $P[d\nu](0.9e^{i\theta})$, $P[d\mu_5](0.9e^{i\theta})$, along with bounds on $P[d\mu](0.9e^{i\theta})$ when $\mu \in \mathcal{F}_{\mathbf{c}_{0:5}}$. It is seen that the spectrum $d\mu_5$ does not distinguish the two peaks. In order to distinguish the two spectral lines, the information in $\mathbf{c}_{0:5}$ is clearly not sufficient as the δ_K -bounds are substantial.

Figure 4 now compares $d\mu_{20}$ and $d\nu$ in a similar manner. The estimation error and the uncertainty diameter are

$$\delta_K(d\nu, d\mu_{20}) = 0.29, \quad \rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:20}}) = 2.52.$$

Here, $d\mu_{20}$ has two peaks close to the spectral lines and $P[d\mu_{20}](0.9e^{i\theta})$ resembles $P[d\nu](0.9e^{i\theta})$ quite closely. In fact, the bounds/envelops already reflect the presence of the two peaks.

To amplify the point made above, consider $d\mu_{\text{line}}$ to be the (unique) power spectrum in $\mathcal{F}_{\mathbf{c}_{0:20}}$ having Schur parameter

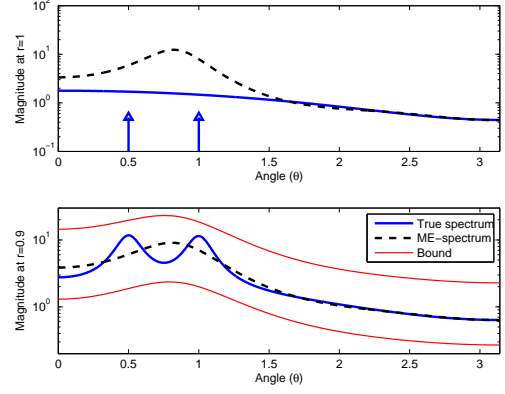


Fig. 3. Subplot 1: The power spectrum $d\nu$ (solid), $d\mu_5$ (dashed). Subplot 2: $P[d\mu](0.9e^{i\theta})$ (solid), $P[d\mu_5](0.9e^{i\theta})$ (dashed), along with bounds based on $\mathbf{c}_{0:5}$.

$\gamma_{21} = 1$; this corresponds to a deterministic process (having only spectral lines) and is depicted in Figure 5. Subplot 2 shows $P[d\mu_{\text{line}}](0.9e^{i\theta})$ and how it “sits” within the respective bounds. In the absence of additional information, $d\mu_{\text{line}}$, $d\mu_{20}$, or any other power spectrum in $\mathcal{F}_{\mathbf{c}_{0:20}}$ is admissible. The “worst case” distance between any two is the diameter computed above.

Remark 15: Even though the three spectra $d\nu$, $d\mu_{20}$, and $d\mu_{\text{line}}$ have identical covariances $\mathbf{c}_{0:20}$, they are quite different in terms of their respective singular and continuous parts. However, they are similar in their distribution of spectral-mass – they have most of their mass located around the frequency points $\theta = 0.5$ and $\theta = 1$, and this is what the weak topology captures.

Remark 16: Standard pointwise distances between $d\nu$, $d\mu_{20}$, and $d\mu_{\text{line}}$ do not provide a meaningful comparison. For instance, the Itakura-Saito distance [23], the Kullback-Leibler divergence [20], and the Cepstral distance [24], because they contain a logarithmic term, give the value of ∞ when comparing $d\mu_{20}$ and $d\mu_{\text{line}}$. On the other hand, the L_2 metric does not apply to the present context because spectral lines cannot be viewed as “ L_2 functions” and if approximated the norm diverges to infinity. Finally, the total variation does not differentiate when spectral lines are nearby or far apart (c.f., Remark 5).

VII. GENERALIZED STATISTICS

Our analysis extends readily to the case of generalized statistics [7], [15], [6], [14]. The formalism in these references, nicknamed THREE (for “tunable high resolution estimation”) allows for the possibility of tunable filter-banks and was shown to provide improved resolution, albeit, quantitative assessments of the benefits exist only in special cases [2]. We briefly sketch the formalism here, for lack of space, and we refer to the aforementioned references for more detailed accounts.

We explain the formalism of generalized statistics in the setting of “filter-banks”, i.e., we consider the stochastic pro-

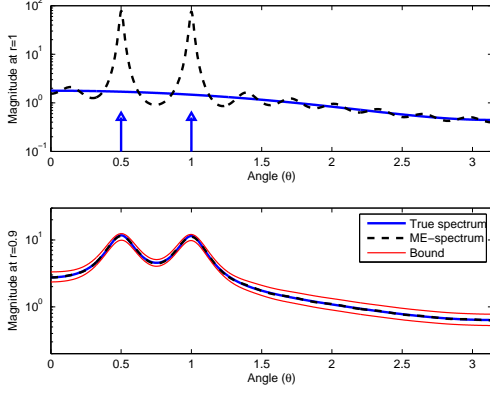


Fig. 4. Subplot 1: Power spectrum $d\nu$ (solid), $d\mu_{20}$ (dashed). Subplot 2: $P[d\mu](0.9e^{i\theta})$ for the true spectrum (solid), $P[d\mu_{20}](0.9e^{i\theta})$ for $d\mu_{20}$ (dashed), along with bounds based on $\mathbf{c}_{0:20}$.

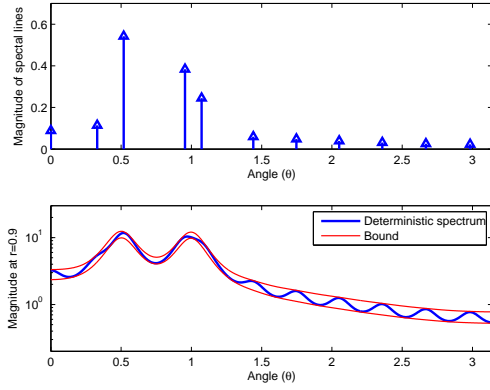


Fig. 5. Subplot 1: Power spectrum $d\mu_{\text{line}}$ (line spectrum). Subplot 2: $P[d\mu_{\text{line}}](0.9e^{i\theta})$ for the line spectrum, along with bounds based on $\mathbf{c}_{0:20}$.

cess y_t as driving a bank of first-order dynamical systems with transfer functions

$$G_k(z) := \frac{z}{z - z_k}, \text{ for } k = 0, 1, \dots, n, \text{ with } |z_k| < 1$$

as shown in Figure 6. The joint covariance matrix of the filter-bank outputs is

$$P = E\{\mathbf{u}(t)\mathbf{u}(t)^*\},$$

where $\mathbf{u} := (u_0(t), u_1(t), \dots, u_n(t))^T$. As indicated earlier $t \in \mathbb{Z}$ is the time index. The covariance matrix takes the form

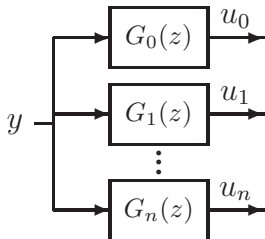


Fig. 6. Bank of filters.

of a Pick matrix

$$P := \left[\frac{w_k + \bar{w}_\ell}{1 - z_k \bar{z}_\ell} \right]_{k, \ell=0}^n \quad (15)$$

where

$$w_k = \frac{1}{2}(1 - z_k^2) \mathcal{E}\{u_k^2\}$$

(see [7, Equations (2.8), (2.10)] and [15, page 783, Equation (7)]). The matrix P replaces the ordinary Toeplitz covariance in the previous sections. Certain observations are in place: given the filter-bank dynamics, i.e., the z_k 's, i) P depends only on the values w_k , and ii) the cross-covariances between filter-bank elements can be computed from the output covariances of all elements individually, that is, from the w_k 's.

A rather complete theory has been developed to characterize power spectra for the input process that are consistent with output-covariance (more generally, state-covariance) statistics. This theory provides among other things a construction of the unique input spectrum of maximal entropy, spectral envelopes that are reminiscent of the Capon pseudo-spectra, and the identification of spectral lines with techniques analogous to the theory of the Pisarenko Harmonic Decomposition, MUSIC, ESPRIT, etc., and has been worked out in detail for matrix-valued power spectra as well (see e.g., [15], [16], [17], [18], [43], [44]).

We restrict our attention to the present setting where $\{y_t\}_{t \in \mathbb{Z}}$ is scalar as before and so are the filters. We assume estimates for the output covariances, hence, the values w_k 's. Like before, we now denote by $\mathcal{F}_{\mathbf{z}, \mathbf{w}}$ the family of power spectra for the process $\{y_t\}_{t \in \mathbb{Z}}$ which are consistent with these values and we are interested in assessing the size of this family as a measure of our spectral uncertainty.

The following proposition can be derived almost verbatim as Proposition 12. See [32] for an independent proof.

Proposition 17: Let z_0, \dots, z_n and w_0, \dots, w_n be such that the Pick matrix P in (15) is positive and let $K \subset \mathbb{D}$ be closed. Then

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{z}, \mathbf{w}}) = \max_{z \in K} \left\{ 2 \left(\left| \frac{\frac{1}{1-z\bar{z}} + \langle b_z, d_z \rangle_{P^{-1}}}{\langle b_z, b_z \rangle_{P^{-1}}} \right|^2 - \frac{\langle d_z, d_z \rangle_{P^{-1}}}{\langle b_z, b_z \rangle_{P^{-1}}} \right)^{\frac{1}{2}} \right\},$$

where

$$b_z = \begin{pmatrix} \frac{1}{1-z_0\bar{z}} \\ \frac{1}{1-z_1\bar{z}} \\ \vdots \\ \frac{1}{1-z_n\bar{z}} \end{pmatrix}, \quad d_z = - \begin{pmatrix} \frac{w_0}{1-z_0\bar{z}} \\ \frac{w_1}{1-z_1\bar{z}} \\ \vdots \\ \frac{w_n}{1-z_n\bar{z}} \end{pmatrix},$$

and $\langle x, y \rangle_{P^{-1}}$ denote the inner product

$$\langle x, y \rangle_{P^{-1}} := y^* P^{-1} x.$$

As before, $\rho_{\delta_K}(\mathcal{F}_{\mathbf{z}, \mathbf{w}})$ is attained as the distance between two elements of $\mathcal{F}_{\mathbf{z}, \mathbf{w}}$ which are both singular with support containing at most $n+1$ points.

As in the covariance case *a priori* bounds on the uncertainty may be calculated.

Theorem 18: Following the notation of Proposition 17, let $z_0 = 0$ and

$$B_{\mathbf{z}}(z) = \prod_{k=0}^n \frac{z - z_k}{1 - \bar{z}_k z}.$$

Then

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{z}, \mathbf{w}}) \leq \max_{z \in K} \frac{4w_0 |B_{\mathbf{z}}(z)|}{1 - |z|^2}. \quad (16)$$

Further, (16) holds with equality if and only if

$$w_k = w_0(1 + z_k \bar{\alpha}) / (1 - z_k \bar{\alpha}) \text{ for } k = 1, \dots, n$$

for some $\alpha \in K$ maximizing $|B_{\mathbf{z}}(\alpha)| / (1 - |\alpha|^2)$.

Proof: The proof is given in the appendix. ■

Here the *a priori* bound depends on the interpolation points \mathbf{z} , in addition to K , the model order n , and the total spectral mass w_0 . Therefore, by minimizing the right hand side of (16) with respect to the \mathbf{z} , one can find the filter bank with the smallest *a priori* uncertainty in the metric δ_K . This will be exploited in the following example to tune the filter-bank poles.

VIII. UNCERTAINTY IN THE THREE FRAMEWORK WITH OPTIMAL FILTER SELECTION

From this vantage point we now take up an example as before, with closely spaced sinusoids, and compare two alternative formalisms, one based on Toeplitz covariances and the other based on generalized statistics.

Consider the stochastic process

$$y_t = \frac{\cos(0.5t + \varphi_1) + \cos(0.6t + \varphi_2)}{2} + \cos(t + \varphi_3) + w_t + \frac{1}{3}w_{t-1},$$

with two closely-spaced spectral lines at 0.5 rad/s and 0.6 rad/s superimposed with a spectral line in 1 and colored noise. We choose as metric δ_K , with $K \subset \mathbb{D}$ proximal to the region where high resolution is desired – i.e., near 0.5 rad/s where the two closely-spaced sinusoids reside. More specifically, we take⁴

$$K := \{0.65e^{\pm 0.5i} + 0.25\mathbb{T}\}.$$

This is depicted by the two circles in Figure 8.

We compare the maximum entropy spectral estimate $d\mu_{\text{ME}}$ constructed using the covariances c_0, c_1, \dots, c_{20} , with the spectral estimate $d\mu_{\text{THREE}}$ which is based on the output statistics of the filter bank of $G_k(z)$'s. We select $n = 10$ and filter-bank poles that minimize⁵ the *a priori* uncertainty bound (16). The filter poles, indicated by “ \times ” in Figure 8, are

$$z_k \in \{0, 0.581 \pm 0.480i, 0.681 \pm 0.470i, 0.738 \pm 0.422i, 0.755 \pm 0.271i, 0.765 \pm 0.357i\}.$$

The THREE-spectrum is a “maximum entropy” distribution which is now consistent with statistics other than the usual autocorrelation ones ($d\mu_{\text{THREE}}$ is the so called “central solution” of the Nevanlinna-Pick analytic interpolation theory⁶ to distributions in $\mathcal{F}_{\mathbf{z}, \mathbf{w}}$).

⁴ \mathbb{T} denotes as before the unit circle.

⁵The pole $z_0 = 0$ and total spectral mass $w_0 = 1$ are assumed fixed. The bound in (16) is then minimized over z_1, \dots, z_n . Since RHS of (16) is nonconvex in \mathbf{z} only local minimum is guaranteed.

⁶Software is available at

http://www.ece.umn.edu/~georgiou/code/spec_analysis.tar

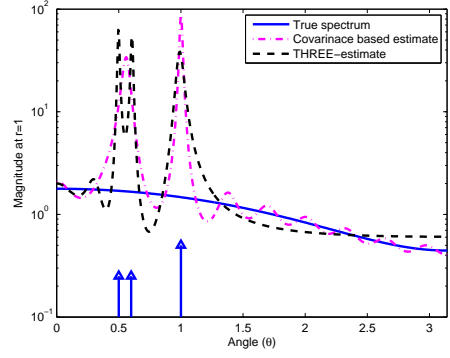


Fig. 7. True spectrum $d\nu$ (solid) and estimated spectra $d\mu_{\text{THREE}}$ (dashed-dotted) and $d\mu_{\text{ME}}$ (dashed).

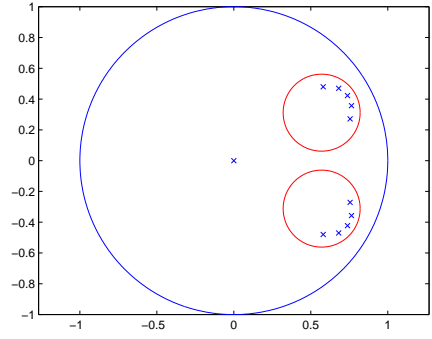


Fig. 8. Set K (solid red) and points z_k (\times in blue).

The *a priori* bounds on the uncertainty provided by Theorems 13 and 18 are

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{z}, \mathbf{w}}) \leq 0.151w_0 = 0.468 \text{ and } \rho_{\delta_K}(\mathcal{F}_{c_{0:n}}) \leq 2.304c_0 = 7.167,$$

respectively. In our example $w_0 = c_0 = 28/9$. This shows that the *a priori* bound on the uncertainty set with respect to δ_K is considerably smaller when the THREE formalism is applied.

The two spectral estimates together with the true power spectrum are depicted in Figure 7. It can be seen that the two closely-spaced lines are not discernible in $d\mu_{\text{ME}}$. On the other hand, they are quite clearly distinguishable via THREE. This is due to the choice of the dynamics \mathbf{z} . As can be seen from the figure, the resolution of $d\mu_{\text{THREE}}$ is substantially higher than that of $d\mu_{\text{ME}}$ in the vicinity of 0.5 rad/s. We would also like to compare the size of the uncertainty set for the two scenarios. The size of the respective diameters are

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{z}, \mathbf{w}}) = 0.194 \text{ and } \rho_{\delta_K}(\mathcal{F}_{c_{0:n}}) = 2.831.$$

Thus, when measured using δ_K , the uncertainty set using the THREE formalism is considerably smaller. Figure 9 displays the Poisson integral of the true power spectrum evaluate on K , and the corresponding bounds⁷.

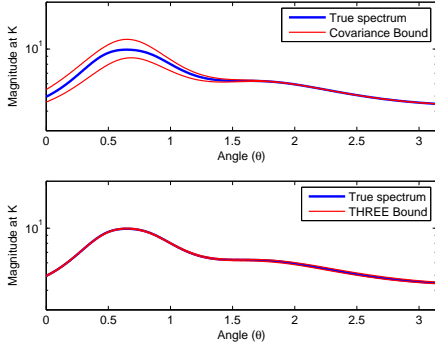


Fig. 9. Bounds on estimates on K based on covariances (top) and the THREE formalism (bottom), respectively.

IX. CONCLUSIONS AND FUTURE DIRECTIONS

The choice of a metric is key to any quantitative scientific theory. Identification of power spectra is often based on second-order statistics (moments), and therefore, it is natural to metrize the space of power spectra in a way that respects continuity of moments. There is a variety of such weakly continuous metrics —metrics which localize “spectral mass”. We presented various choices and focused on a particular metric, δ_K , which is amenable to quantifying the size of the uncertainty set. We envision that this, and similar metrics, can be used as tools for assessing uncertainty and robustness in modeling and spectral analysis. We further expect that the theory will be of use in filter design and in quantifying the notion of resolution—as this is naturally connected to the size of the spectral uncertainty set. Finally, we expect that these metrics will conform with other subjective measures rooted in perceptual qualities of signals (cf. [19, Example 10]).

Interest in weak continuity is not new. Indeed, a classical weakly continuous metric is the Lévy-Prokhorov metric [41] and it is well known that the periodogram converges weakly as the sample size goes to infinity (see, e.g., [39]). Yet, appropriate weakly continuous metrics that can be used to quantify uncertainty have not received much attention —the commonly used “total variation,” Itakura-Saito, and other distance measures are not weakly continuous. Besides the relevance in uncertainty quantification and in filter design (cf. Section VIII), computationally amenable and easy-to-use metrics may provide a useful geometric setting for modeling slowly time-varying processes and for integrating data from disparate sources (see, e.g., [28], [29], [30], [46], [47]).

APPENDIX

Proof: [Theorem 1]

The canonical neighborhood basis for a point $d\nu$ in the weak topology on \mathfrak{M} consists of sets of the type

$$\begin{aligned} N(d\nu, \{g_k\}_{k=1}^n, \epsilon) \\ = \left\{ d\mu \geq 0 : \left| \int_{\mathbb{T}} g_k(d\nu - d\mu) \right| < \epsilon, k = 0, 1, \dots, n \right\}, \end{aligned}$$

⁷Since K is formed out of two circles symmetrically located with respect to the real axis of the complex plane, plots are identical for the two components of K .

where g_k are continuous functions on \mathbb{T} for $k = 0, \dots, n$. To establish the theorem we prove that the neighbourhood basis

$$\mathfrak{N}(d\nu) = \{N(d\nu, \{g_k\}_{k=0}^n, \epsilon) : \epsilon > 0, n \in \mathbb{N}, \{g_k\}_{k=0}^n \subset C(\mathbb{T})\}$$

is equivalent to the basis

$$\mathfrak{F}(d\nu) = \left\{ \mathcal{F}_{\mathbf{c}_{0:n}, \epsilon} : \epsilon > 0, n \in \mathbb{N}, c_k = \int_{\mathbb{T}} z^{-k} d\nu, k = 0, \dots, n \right\}.$$

First note that $\mathfrak{N}(d\nu) \supset \mathfrak{F}(d\nu)$, and hence the weak topology is at least as strong as the topology induced by $\mathfrak{F}(d\nu)$. To establish the other direction, let N be an arbitrary set in $\mathfrak{N}(d\nu)$. To show the equivalence, it is enough to show that there exists $n \in \mathbb{N}$ such that $\mathcal{F}_{\mathbf{c}_{0:n}, \epsilon} \subset N$.

Let δ be a weakly continuous metric for \mathfrak{M} and choose ϵ so that the $B_\delta(d\nu, \epsilon) = \{d\mu \geq 0 : \delta(d\mu, d\nu) < \epsilon\} \subset N$. Next, take $d\mu_\ell \in \mathcal{F}_{\mathbf{c}_{0:\ell}, \ell-1}$ with

$$\delta(d\mu_\ell, d\nu) \geq \frac{1}{2} \sup\{\delta(d\mu, d\nu) : d\mu \in \mathcal{F}_{\mathbf{c}_{0:\ell}, \ell-1}\} \quad (17)$$

for $\ell \geq 1$. Since $\mathcal{F}_{\mathbf{c}_{0:\ell}, \ell-1} \subset \{d\mu : \mu(\mathbb{T}) \leq \nu(\mathbb{T}) + 1\}$, which is weakly compact, there is a convergent subsequence of $d\mu_k$ that converges to $d\hat{\mu}$ (by Banach-Alaoglu [45]). Note that $\mathcal{F}_{\mathbf{c}_{0:\ell}, \ell-1} \supset \text{closure}(\mathcal{F}_{\mathbf{c}_{0:\ell+1}, (\ell+1)-1}) \supset \mathcal{F}_{\mathbf{c}_{0:\ell+1}, (\ell+1)-1}$, hence $d\hat{\mu} \in \mathcal{F}_{\mathbf{c}_{0:\ell}, \ell-1}$ for any ℓ . It then follows that $d\hat{\mu} = d\nu$ since the trigonometric moment problem is determinate (by Riesz-Herglotz, see [1]). Let n be such that $\delta(d\mu_n, d\nu) < \epsilon/2$, then by (17) we have that $\mathcal{F}_{\mathbf{c}_{0:n}, n-1} \subset B_\delta(d\nu, \epsilon) \subset N$. We have thus shown that the topology induced by the neighbourhood basis $\mathfrak{F}(d\nu)$ is the weak topology, and hence δ is weakly continuous if and only if (4) holds. ■

Proof: [Proposition 7]

It is clear that condition (a) holds if and only if $\delta(d\mu_0, d\mu_1)$ is positive whenever $d\mu_0 \neq d\mu_1$. The triangle inequality and symmetry always holds for such δ , so we only need to show that condition (b) holds if and only if δ is weakly continuous.

We will show that condition (b) implies that δ is weakly continuous by contradiction. Assume therefore that condition (b) holds, but that δ is not weakly continuous. Then there exists $d\mu_k \rightarrow d\mu$ weakly such that $\delta(d\mu_k, d\mu) > \epsilon$, $k = 1, 2, \dots$, and hence there exists $g_{\xi_k}, \xi_k \in K$, such that

$$\epsilon < \left| \int_{\mathbb{T}} g_{\xi_k}(d\mu_k - d\mu) \right|, \quad k = 1, 2, \dots$$

To this end we use the Arzelà-Ascoli theorem (see e.g., [33, page 102]) which states that a set of functions is relatively compact⁸ in $C(\mathbb{T})$ if and only if the set of functions is uniformly bounded and equicontinuous. Therefore, since (b) holds, the set $\{g_\xi\}_{\xi \in K}$ is relatively compact in $C(\mathbb{T})$, and there is a subsequence $(g_\ell, d\mu_\ell)$ of $(g_{\xi_k}, d\mu_k)$ such that $g_\ell \rightarrow g \in C(\mathbb{T})$. A contradiction follows, since

$$\begin{aligned} \epsilon &< \left| \int_{\mathbb{T}} g_\ell(d\mu_\ell - d\mu) \right| \\ &\leq \|g_\ell - g\|_\infty \int_{\mathbb{T}} |d\mu_\ell - d\mu| + \left| \int_{\mathbb{T}} g(d\mu_\ell - d\mu) \right| \\ &\rightarrow 0 \text{ as } \ell \rightarrow \infty, \end{aligned}$$

and hence δ is weakly continuous whenever condition (b) holds.

⁸A set is relatively compact if its closure is compact.

Next, we show that (b) holds if δ is weakly continuous, and once again we use contradiction. That is, we show that if (b) fails to be true then δ is not weakly continuous. If $\{g_\xi\}_{\xi \in K}$ is not equicontinuous, then there exists an $\epsilon > 0$ such that for any $k = 1, 2, \dots$ one can find $\theta_k, \phi_k \in \mathbb{T}$, and $\xi_k \in K$, that satisfies

$$|\theta_k - \phi_k| < \frac{1}{k} \text{ and } |g_{\xi_k}(\theta_k) - g_{\xi_k}(\phi_k)| > \epsilon. \quad (18)$$

Let (θ_ℓ, ϕ_ℓ) be a subsequence of (θ_k, ϕ_k) such that $\theta_\ell \rightarrow \theta_0 \in \mathbb{T}$ as $\ell \rightarrow \infty$, and let $d\mu_\ell$ and $d\nu_\ell$ be the measures that consist of a unit mass in θ_ℓ and ϕ_ℓ , respectively. From (18) it follows that $\phi_\ell \rightarrow \theta_0$, and hence that $d\mu_\ell \rightarrow d\mu_0$ and $d\nu_\ell \rightarrow d\mu_0$ weakly, where $d\mu_0$ is the measure that consists of a unit mass in θ_0 . From (18) it follows that

$$\begin{aligned} \delta(d\mu_\ell, d\mu_0) + \delta(d\nu_\ell, d\mu_0) &\geq \delta(d\mu_\ell, d\nu_\ell) \\ &\geq |g_{\xi_\ell}(\theta_\ell) - g_{\xi_\ell}(\phi_\ell)| > \epsilon. \end{aligned}$$

From this, it is evident that δ is not weakly continuous since both $\delta(d\mu_\ell, d\mu_0)$ and $\delta(d\nu_\ell, d\mu_0)$ cannot converge to 0.

Similarly, if $\{g_\xi\}_{\xi \in K}$ is not uniformly bounded, then for any $k = 1, 2, \dots$ one can find $\theta_k \in \mathbb{T}$ and $\xi_k \in K$ such that

$$|g_{\xi_k}(\theta_k)| > k. \quad (19)$$

Let $d\mu_k$ be the measures that consist of a unit mass in θ_k . Therefore, the metric δ is not weakly continuous since $\frac{1}{k}d\mu_k \rightarrow 0$ weakly, while $\delta(\frac{1}{k}d\mu_k, 0) > 1$ for all k . ■

Proof: [Proposition 10]

(a) \Rightarrow (b) $\mu_k \rightarrow \mu$ weakly is equivalent to $\int_{-\pi}^{\pi} f(t)d\mu_k(t) \rightarrow \int_{-\pi}^{\pi} f(t)d\mu(t)$ for all periodic continuous functions $f(t)$. For all $z = re^{i\theta} \in \mathbb{D}$, $P_r(\theta - t)$ is periodic and continuous, hence

$$\begin{aligned} u_k(z) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - t)d\mu_k(t) \\ &\rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - t)d\mu(t) = u(z). \end{aligned}$$

(b) \Rightarrow (c). For $r < 1$, $|u_k(re^{i\theta})| \leq \frac{1+r}{1-r}|\mu_k|(\mathbb{T})$. Since $u_k(re^{i\theta}) \rightarrow u(re^{i\theta})$ pointwise for all θ , it follows from bounded convergence that $\int_{-\pi}^{\pi} |u_k(re^{i\theta}) - u(re^{i\theta})|d\theta \rightarrow 0$. Furthermore, $\forall k, r$, $\int_{-\pi}^{\pi} |u_k(re^{i\theta}) - u(re^{i\theta})|d\theta \leq 2\pi(|\mu_k|(\mathbb{T}) + |\mu|(\mathbb{T}))$ which is uniformly bounded, hence

$$\int_0^1 \int_{-\pi}^{\pi} |u_k(re^{i\theta}) - u(re^{i\theta})|d\theta r dr \rightarrow 0$$

by dominated convergence.

(c) \Rightarrow (d). Let $K \subset \mathbb{D}$ be a compact set. Then there exist an $\epsilon > 0$ such that $B_\epsilon(z_0) = \{z : |z - z_0| < \epsilon\} \subset \mathbb{D}$ for all $z_0 \in K$. Now by the mean value property of harmonic functions we have

$$\begin{aligned} u(z_0) &= \frac{2}{\epsilon^2} \int_0^\epsilon u(z_0)r dr \\ &= \frac{2}{\epsilon^2} \int_0^\epsilon \frac{1}{2\pi} \int_{-\pi}^{\pi} u(z_0 + re^{i\theta})d\theta r dr \\ &= \frac{1}{\pi\epsilon^2} \int_{B_\epsilon(z_0)} u(z)dx dy. \end{aligned}$$

Of course the same equality holds for $u_k(z_0)$

$$u_k(z_0) = \frac{1}{\pi\epsilon^2} \int_{B_\epsilon(z_0)} u_k(z)dx dy.$$

For any $z_0 \in K$ the difference between the harmonic functions is bounded by

$$\begin{aligned} |u_k(z_0) - u(z_0)| &\leq \frac{1}{\pi\epsilon^2} \int_{B_\epsilon(z_0)} |u_k(z) - u(z)|dx dy \\ &\leq \frac{1}{\pi\epsilon^2} \int_{\mathbb{D}} |u_k(z) - u(z)|dx dy. \end{aligned}$$

By (c) the difference goes to zero uniformly in K .

(d) \Rightarrow (a). Let $f \in C(\mathbb{T})$. For any bounded measure $\nu \in \mathcal{F}$ and corresponding harmonic function $v(z) = P[\nu](z)$ Fubini's theorem gives

$$\begin{aligned} \int_{-\pi}^{\pi} f(t)v(re^{it})dt &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \int_{-\pi}^{\pi} P_r(\theta - t)f(t)dt d\nu(\theta) \\ &= \int_{-\pi}^{\pi} P[f(t)dt](re^{i\theta})d\nu(\theta). \end{aligned}$$

Since f is periodic and continuous, $P[f(t)dt](re^{i\theta})$ converges uniformly to $f(\theta)$, hence

$$\begin{aligned} \left| \int_{-\pi}^{\pi} f(t)v(re^{it})dt - \int_{-\pi}^{\pi} f(t)d\nu(t) \right| &\leq \\ \|P[f](re^{it}) - f(t)\|_\infty |\nu|(\mathbb{T}) \end{aligned}$$

converges to zero independent of the measure ν . This shows that for an arbitrary $\epsilon > 0$ there exists an $0 < r < 1$ such that

$$\left| \int_{-\pi}^{\pi} f(t)v(re^{it})dt - \int_{-\pi}^{\pi} f(t)d\nu(t) \right| < \frac{\epsilon}{3}$$

for $\nu \in \{\mu, \mu_1, \mu_2, \dots\}$. Further more, since $u_k \rightarrow u$ uniformly on $\{z : |z| \leq r\}$, it is possible to find an $k_{r,\epsilon}$ be such that

$$\left| \int_{-\pi}^{\pi} f(t)u_k(re^{it})dt - \int_{-\pi}^{\pi} f(t)u(re^{it})dt \right| < \frac{\epsilon}{3}$$

for all $k > k_{r,\epsilon}$. By the triangle inequality we have

$$\left| \int_{-\pi}^{\pi} f(t)d\mu_k(t) - \int_{-\pi}^{\pi} f(t)d\mu(t) \right| < \epsilon$$

for all $k > k_{r,\epsilon}$. Since ϵ was chosen arbitrarily, $\left| \int_{-\pi}^{\pi} f(t)d\mu_k(t) - \int_{-\pi}^{\pi} f(t)d\mu(t) \right| \rightarrow 0$ as $k \rightarrow \infty$, and weak convergence follows. ■

Proof: [Proposition 12]

There exists an analytic function $f(z) = H[d\mu](z)$, $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$, such that $f(z) = w_z$ if and only if its associated Pick matrix is nonnegative [34], i.e.

$$\begin{pmatrix} 2T_n & b_z w_z - d_z \\ \bar{w}_z b_z^* - d_z^* & \frac{w_z + \bar{w}_z}{1 - z\bar{z}} \end{pmatrix} \geq 0. \quad (20)$$

By using Schur's lemma and completing the square, we arrive at

$$\begin{aligned} &\left| w_z - \frac{\frac{2}{1-z\bar{z}} + \langle d_z, b_z \rangle_{T^{-1}}}{\langle b_z, b_z \rangle_{T^{-1}}} \right|^2 \\ &\leq \left| \frac{\frac{2}{1-z\bar{z}} + \langle b_z, d_z \rangle_{T^{-1}}}{\langle b_z, b_z \rangle_{T^{-1}}} \right|^2 - \frac{\langle d_z, d_z \rangle_{T^{-1}}}{\langle b_z, b_z \rangle_{T^{-1}}}, \end{aligned} \quad (21)$$

where equality holds if and only if the Pick matrix (20) is singular. From this, the first part of Proposition 12 follows. Since the maximum is obtained when equality holds in (21),

the associated Pick matrices are singular. Hence the solutions are unique and correspond to measures with support on $n+1$ points [15, Proposition 2]. ■

Background on orthogonal polynomials and Schur coefficients

Let \mathbf{c} be a nonnegative covariance sequence with corresponding measure $d\mu$ and consider the inner product

$$\langle a(z), b(z) \rangle = \frac{1}{2\pi} \int_{\mathbb{T}} a(e^{i\theta}) \overline{b(e^{i\theta})} d\mu(\theta).$$

The so-called *orthogonal polynomials (of the first kind)* $\phi_k(z)$ [22] are (uniquely defined) monic polynomials with $\deg \phi_k(z) = k$, $k = 0, 1, \dots$, which are orthogonal with respect to $\langle \cdot, \cdot \rangle$. They are shown [22] to satisfy the recursion

$$\begin{aligned} \phi_{k+1}(z) &= z\phi_k(z) - \bar{\gamma}_k \phi_k(z)^*, \\ \phi_{k+1}(z)^* &= \phi_k(z)^* - z\gamma_k \phi_k(z), \end{aligned} \quad (22)$$

where $\phi_k(z)^* = z^k \overline{\phi_k(\bar{z}^{-1})}$ and $\{\gamma_k\}_{k=1}^\infty$ are the so-called *Schur parameters*.

The *orthogonal polynomials of the second kind* are defined by

$$\psi_k(z) = \frac{1}{c_0} [(f(\bar{z}^{-1}))\phi_k(z)]_+,$$

where $[\cdot]_+$ denote “the polynomial part of”. They are also “orthogonal polynomials” but with respect to a certain “inverted” covariance (corresponding to the negative of the original Schur parameters, cf. [22]) and satisfy the recursion

$$\begin{aligned} \psi_{k+1}(z) &= z\psi_k(z) + \bar{\gamma}_k \psi_k(z)^*, \\ \psi_{k+1}(z)^* &= \psi_k(z)^* + z\gamma_k \psi_k(z). \end{aligned} \quad (23)$$

The positive-real function $f(z) = H[d\mu](z)$ may be expressed using the orthogonal polynomials as

$$f(z) = c_0 \frac{\psi_k(z)^* + z s_{k+1}(z) \psi_k(z)}{\phi_k(z)^* - z s_{k+1}(z) \phi_k(z)}, \quad (24)$$

where $s_{k+1}(z)$ belong to the Schur class \mathcal{S} , i.e. the class of analytic functions on \mathbb{D} uniformly bounded by 1. Equations (22-23) lead to

$$f(z) = c_0 \frac{1 + z s_1(z)}{1 - z s_1(z)}, \quad s_k(z) = \frac{\gamma_k + z s_{k+1}(z)}{1 + z \bar{\gamma}_k s_{k+1}(z)}, \quad (25)$$

for $k = 1, 2, \dots$. For a complete exposition on orthogonal polynomials and Schur’s algorithm see [1], [22], [25].

Proof: [Theorems 13 and 18]

Following our earlier notation, let

$$\rho_{\delta_\alpha}(\mathcal{F}_{\mathbf{c}_{0:n}}) = \max\{ |P[d\mu_0](\alpha) - P[d\mu_1](\alpha)| : d\mu_0, d\mu_1 \in \mathcal{F}_{\mathbf{c}_{0:n}} \}$$

be the uncertainty diameter at the point $\alpha \in K$. By using (24) and noting that

$$P[d\mu](\alpha) = \Re e H[d\mu](\alpha) = \Re e f(\alpha),$$

the diameter $\rho_{\delta_\alpha}(\mathcal{F}_{\mathbf{c}_{0:n}})$ is equal to the diameter of the disc

$$c_0 \frac{\psi_n(\alpha)^* + \alpha s_{n+1}(\alpha) \psi_n(\alpha)}{\phi_n(\alpha)^* - \alpha s_{n+1}(\alpha) \phi_n(\alpha)} : s_{n+1}(\alpha) \in \overline{\mathbb{D}}, \quad (26)$$

where ϕ_n and ψ_n are specified via (22), (23), by the Schur sequence $(\gamma_1, \dots, \gamma_n)$ corresponding to $\mathbf{c}_{0:n}$ (see [25]). Denote by r_n the radius of (26), and hence $\rho_{\delta_\alpha}(\mathcal{F}_{\mathbf{c}_{0:n}}) = 2r_n$.

For $n = 0$, we have $\phi_0(z) = \psi_0(z) = 1$. The expression in (26) is

$$c_0 \frac{1 + \alpha s_1}{1 - \alpha s_1} = c_0 \frac{1 + |\alpha|^2}{1 - |\alpha|^2} + \frac{2c_0\alpha}{1 - |\alpha|^2} \frac{-\bar{\alpha} + s_1}{1 - \alpha s_1},$$

hence $r_0 = 2c_0\alpha/(1 - |\alpha|^2)$, where s_k without argument denotes $s_k(\alpha)$. Next, consider the radius of (26) for $n = k - 1$. The set (26) is the range of a Möbius transform applied to $s_k \in \overline{\mathbb{D}}$, and may be represented as

$$M_{k-1} + e^{i\theta_{k-1}} r_{k-1} \frac{v_{k-1} + s_k}{1 + \bar{v}_{k-1} s_k} \quad (27)$$

where M_{k-1} and r_{k-1} are the center and radius of the disc, respectively, and where $\theta_{k-1} \in (-\pi, \pi]$, $v_{k-1} \in \mathbb{D}$. From the recursion (25), it can be seen that

$$\frac{v_{k-1} + s_k}{1 + \bar{v}_{k-1} s_k} = \frac{1 + \bar{\gamma}_k v_{k-1}}{1 + \gamma_k \bar{v}_{k-1}} \frac{\eta_k + \alpha s_{k+1}}{1 + \bar{\eta}_k \alpha s_{k+1}},$$

where

$$\eta_k = \frac{v_{k-1} + \gamma_k}{1 + \bar{\gamma}_k v_{k-1}}.$$

The set (26) for $n = k$ is therefore

$$M_{k-1} + e^{i\theta_{k-1}} \frac{1 + \bar{\gamma}_k v_{k-1}}{1 + \gamma_k \bar{v}_{k-1}} r_{k-1} \frac{\eta_k + \alpha s_{k+1}}{1 + \bar{\eta}_k \alpha s_{k+1}} : s_{k+1} \in \mathbb{D}.$$

A Möbius transformation $(a + bs)/(c + ds)$, with $|c| > |d|$, maps the unit disc to a disc of radius $|ac - bd|/(|c|^2 - |d|^2)$. Therefore, the radius

$$r_k = r_{k-1} |\alpha| \frac{1 - |\eta_k|^2}{1 - |\eta_k|^2 |\alpha|^2} = r_{k-1} |\alpha| \left(1 - \frac{|\eta_k|^2 (1 - |\alpha|^2)}{1 - |\eta_k|^2 |\alpha|^2} \right)$$

is maximized when $\eta_k = 0$, or equivalently when $\gamma_k = -v_{k-1}$. Hence, $r_k \leq r_{k-1} |\alpha|$ with equality if $\gamma_k = -v_{k-1}$. By induction, the maximal radius is given by

$$r_n = r_0 |\alpha|^n = 2c_0 |\alpha|^{n+1} / (1 - |\alpha|^2).$$

Furthermore, $\gamma_k = -v_{k-1}$ in the recursion (25) correspond to the Schur parameters $\gamma_1 = \bar{\alpha}$, and $\gamma_k = 0$ for $k = 2, \dots, n$. This leads to the covariance sequence $\mathbf{c}_{0:n} = (c_0, c_0 \bar{\alpha}, \dots, c_0 \bar{\alpha}^n)$. Since δ_K is defined as the maximal diameter over all $\alpha \in K$, the inequality

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}}) = \max_{\alpha \in K} \rho_{\delta_\alpha}(\mathcal{F}_{\mathbf{c}_{0:n}}) \leq \max_{\alpha \in K} \frac{4|\alpha|^{n+1} c_0}{1 - |\alpha|^2}$$

holds and is achieved for $\mathbf{c}_{0:n} = (c_0, c_0 \bar{\alpha}, \dots, c_0 \bar{\alpha}^n)$ where $\alpha \in K$ maximizes $|\alpha|^{n+1}/(1 - |\alpha|^2)$.

In the Nevalinna-Pick case, the recursion is identical to (25) except that z is replaced by the inner factor $\xi_k(z) = (z_k - z)/(1 - \bar{z}_k z)$

$$f(z) = w_0 \frac{1 + z s_1(z)}{1 - z s_1(z)}, \quad s_k(z) = \frac{\gamma_k + \xi_k(z) s_{k+1}(z)}{1 + \bar{\gamma}_k \xi_k(z) s_{k+1}(z)},$$

for $k = 1, 2, \dots, n$ [13]. The argument here is analogous to the covariance case. The shrinkage of the radius is $r_k \leq r_{k-1} |\xi_k(\alpha)|$, with equality when the parameters in the recursion are $(\bar{\alpha}, 0, \dots, 0)$, as in the covariance case. The bound of the uncertainty diameter at α then becomes

$$2r_n \leq \frac{4w_0 |\alpha| \prod_{k=1}^n |\xi_k(\alpha)|}{1 - |\alpha|^2} = \frac{4w_0 |B_{\mathbf{z}}(\alpha)|}{1 - |\alpha|^2} \quad (28)$$

which is attained when $w_k = w_0(1 + z_k \bar{\alpha})/(1 - z_k \bar{\alpha})$ for $k = 1, \dots, n$. Since $\rho_{\delta_\alpha}(\mathcal{F}_{\mathbf{z}, \mathbf{w}}) = 2r_n$, maximizing (28) for $\alpha \in K$ gives the bound (16). ■

REFERENCES

- [1] N. I. Akhiezer, **The classical moment problem**, Hafner Publishing, Translation 1965, Oliver and Boyd.
- [2] A. N. Amini and T. T. Georgiou, "Tunable Line Spectral Estimators Based on State-Covariance Subspace Analysis," *IEEE Trans. on Signal Processing*, **54(7)**: 2662-2671, July 2006.
- [3] E. Avventi, "Spectral Moment Problems: Generalizations, Implementations and Tuning," PhD thesis, KTH, Stockholm, Sweden, 2011.
- [4] A. Ball, I. Gohberg, and L. Rodman, **Interpolation of rational matrix functions**, Operator Theory: Advances and Applications Boston: Birkhäuser, vol. 45, 1990.
- [5] A. Blomqvist, A. Lindquist, and R. Nagamune, "Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: an optimization approach," *IEEE Trans. on Automatic Control*, **48(12)**: 2172-2190, December 2003.
- [6] C. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. on Automatic Control*, **46(6)**: 822-839, June 2001.
- [7] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to Spectral Estimation: A tunable high-resolution spectral estimator," *IEEE Trans. on Signal Processing*, **48(11)**: 3189-3205, November 2000.
- [8] J. Capon, "High-resolution frequency-wave number spectrum analysis," *IEEE Proc.*, **57(8)**: 1408-1418, August 1969.
- [9] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger Versus Kullback-Leibler Multivariable Spectrum Approximation," *IEEE Trans. on Automatic Control*, **53(4)**: 954-967, May 2008.
- [10] A. Ferrante, M. Pavon, and M. Zorzi, "A Maximum Entropy Enhancement for a Family of High-Resolution Spectral Estimators," *IEEE Trans. on Automatic Control*, **57(2)**: 318-329, February 2012.
- [11] A. Ferrante, F. Ramponi, and F. Ticozzi, "On the Convergence of an Efficient Algorithm for Kullback-Leibler Approximation of Spectral Densities," *IEEE Trans. on Automatic Control*, **56(3)**: 506-515, March 2011.
- [12] C. Foias, A. E. Frazho, **The commutant lifting approach to interpolation problems**, Birkhauser Verlag, Ot 44: Advances and Applications, 1990.
- [13] J. B. Garnett, **Bounded Analytic Functions**, Springer, San Diego, 2007.
- [14] T. T. Georgiou, "Spectral Estimation via Selective Harmonic Amplification," *IEEE Trans. on Automatic Control*, **46(1)**: 29-42, January 2001.
- [15] T. T. Georgiou, "Spectral Estimation via Selective Harmonic Amplification: MUSIC, Redux," *IEEE Trans. on Signal Processing*, **48(3)**: 780-790, March 2000.
- [16] T. T. Georgiou, "Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization," *IEEE Trans. on Automatic Control*, **47(11)**: 1811-1823, November 2002.
- [17] T. T. Georgiou, "The Carathéodory-Fejér-Pisarenko decomposition and its multivariable counterpart," *IEEE Trans. on Automatic Control*, **52(2)**: 212-228, February 2007.
- [18] T. T. Georgiou, "Relative Entropy and the multi-variable multi-dimensional Moment Problem," *IEEE Trans. on Information Theory*, **52(3)**: 1052-1066, March 2006.
- [19] T. T. Georgiou, J. Karlsson, and M. S. Takyar, "Metrics for power spectra: an axiomatic approach," *IEEE Trans. on Signal Processing*, **57(3)**: 859-867, March 2009.
- [20] T. T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. on Information Theory*, **49(11)**: 2910-2917, November 2003.
- [21] T. T. Georgiou and A. Lindquist, "A Convex Optimization Approach to ARMA Modeling," *IEEE Trans. on Automatic Control*, **53(5)**: 1108-1119, June 2008.
- [22] Ya. L. Geronimus, **Orthogonal Polynomials**, Consultants Bureau, 210 pages, 1961.
- [23] R. Gray, A. Buzo, A. Gray Jr., and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. on Acoustics, Speech and Signal Processing*, **28(4)**: 367-376, August 1980.
- [24] A. Gray Jr. and J. Markel, "Distance measures for speech processing," *IEEE Trans. on Acoustics, Speech and Signal Processing*, **24(5)**: 380-391, October 1976.
- [25] U. Grenander and G. Szegő, **Toeplitz Forms and their Applications**, Chelsea, 1958.
- [26] S. Haykin, **Nonlinear Methods of Spectral Analysis**, Springer-Verlag, New York, 247 pages, 1979.
- [27] K. Hoffman, **Banach Spaces of Analytic Functions**, Dover Publications, 216 pages, 1962.
- [28] X. Jiang, M. S. Takyar, and T. T. Georgiou, "Metrics and morphing of power spectra," in *Lecture Notes in Control and Information Sciences*, V. Blondel, S. Boyd, H. Kimura eds., **371**, Springer Verlag 2008.
- [29] X. Jiang, J. Karlsson, and T. T. Georgiou, "Phoneme segmentation based on spectral metrics," *International Symposium on Mathematical Theory of Networks and Systems*, Blacksburg, VA, July 2008, (abstract).
- [30] X. Jiang, Z.-Q. Luo, and T. T. Georgiou, "Spectral geodesics and tracking," *Proc. IEEE Conference on Decision and Control*, Cancun, Mexico, December 2008.
- [31] J. Karlsson and T. T. Georgiou, "Signal analysis, moment problems & uncertainty measures," *Proc. IEEE Conference on Decision and Control*, Seville, Spain, December 2005.
- [32] J. Karlsson and T. T. Georgiou, "Metric Uncertainty for Spectral Estimation based on Nevanlinna-Pick Interpolation," *International Symposium on Mathematical Theory of Networks and Systems*, Melbourne, Australia, July 2012.
- [33] A. N. Kolmogorov and S. V. Fomin, **Introductory Real Analysis**, Dover Publications, 1970.
- [34] I. V. Kovalishina and V. P. Potapov, **Integral Representation of Hermitian Positive Functions**, Khark'ov Railway Engineering Inst., Khark'ov 2001. English translation by T. Ando, Sapporo, Japan, 1981.
- [35] M. G. Krein and A. A. Nudel'man, **The Markov Moment Problem and Extremal Problems**, American Mathematical Society, Providence, RI, 417 pages, 1977.
- [36] S. Lang and T. Marzetta, "A linear programming approach to bounding spectral power," *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing*, pp. 847-850, Boston, MA, April 1983.
- [37] T. Marzetta and S. Lang, "Power spectral density bounds," *IEEE Transactions on Information Theory*, **30(1)**: 117-122, January 1984.
- [38] R. Nagamune, "A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. on Automatic Control*, **48(1)**: 113-117, January 2003.
- [39] E. Parzen, "Mathematical Considerations in the estimation of Spectra," *Technometrics*, **3(2)**: 167-190, May 1961.
- [40] M. Pavon and A. Ferrante, "On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities," *IEEE Trans. on Automatic Control*, **51(4)**: 639-644, April 2006.
- [41] Yu. V. Prokhorov, "Convergence of random processes and limit theorems in probability theory," *Theory Probab. Appl.*, **1**: 157-214, 1956.
- [42] S. T. Rachev and L. Rüschendorf, **Mass Transportation Problems: Theory, vols. I and II**, Probability and its Applications. Springer: New York, 1998.
- [43] F. Ramponi, A. Ferrante, and M. Pavon, "A Globally Convergent Matricial Algorithm for Multivariate Spectral Estimation," *IEEE Trans. on Automatic Control*, **54(10)**: 2376-2388, October 2009.
- [44] F. Ramponi, A. Ferrante, and M. Pavon, "On the well-posedness of multivariate spectrum approximation and convergence of high-resolution spectral estimators," *System and Control Letters*, **59(3-4)**: 167-172, March-April 2010.
- [45] W. Rudin, **Functional analysis**, McGraw-Hill, 397 pages, 1973.
- [46] D. Rudoy, "Nonstationary Time Series Modeling with Application to Speech Processing," Ph.D. thesis, Harvard University, December 2010.
- [47] D. Rudoy and T. T. Georgiou, "Regularized Parametric Models of Nonstationary Processes," *International Symposium on Mathematical Theory of Networks and Systems*, Budapest, Hungary, July 2010.
- [48] P. Stoica and R. Moses, **Introduction to Spectral Analysis**, Prentice Hall, 1997.
- [49] P. P. Vaidyanathan, **Multirate Systems and Filter Banks**, Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [50] C. Villani, **Topics in Optimal Transportation**, Graduate studies in Mathematics, vol 58, AMS, 2003.